msCompare

Hoekman, Berend; Breitling, Rainer; Suits, Frank; Bischoff, Rainer; Horvatovich, Peter

Published in:
Molecular & Cellular Proteomics

DOI:
10.1074/mcp.M111.015974

Publication date:
2012

Supplementary material

1. Experimental and Computational methods

1.1. FeatureList and FeatureMatrix XML formats

The various open-source data processing workflows that were integrated into the msCompare framework all use their own internal data formats to transfer information between modules. To allow communication between feature detection/quantification, and alignment/matching modules of the different tools, two generic XML formats for the FeatureList and the FeatureMatrix were developed to store feature lists along with matched quantitative feature matrices (Figure 1a, b). The main aim of the developed formats is to store all information provided by the integrated programs, making conversion between the different formats possible and allowing easy integration of new modules from other programs. The FeatureList XML format (Figure S7 in supplementary material) starts with a *Data* element, which has a *data file* attribute containing the link to the original mzXML file that was used to generate the feature list. The *Data* node has one *FeatureList* child node, which contains one or more *Feature* nodes. Each feature node represents a detected and quantified feature in the raw LC-MS data and has an *AttributeList* and a *HullPointList* child node. The *HullPointList* node may have zero or more *HullPoint* child nodes. The *HullPoint* element has one or more *Attribute* nodes, which have two attributes, value and name. The following names are used within msCompare to define the properties of a HullPoint: *retentionTime* to store the retention time in minutes, *mass* to store the mass to charge ratio value in amu, *scanNumber* to store the scan number, and *intensity* to store the signal associated with this point. The *AttributeList* node has zero or more *Attribute* nodes. The *Attribute* node has three attributes: *programName* indicating the name of the program that was used to generate this attribute, *attributeName* containing the name of the attribute, which is specific for the program and *value* containing the value that is associated with the attribute. The content of the *value* and the *attributeName* attributes can be any text (including numbers), thus allowing to store any kind of information related to a feature. This generic format allows the integration of any new feature list format due to the flexibility provided by defining the feature properties as "values" instead of using fixed XML attributes. The generic FeatureMatrix XML format has a similar structure as the generic feature list format. The XML description begins with a *Data* node containing a *ClassMap* and a *FeatureArray* node. The *ClassMap* node contains, for each feature list file that is used to build the quantitative matched feature matrix, one *Map* child element with two attributes: *filename* containing the name of one of the feature list files as used for the matching process and *classNum* indicating the class identifier of the sample from which the feature list was generated. The *FeatureArray* element contains zero or more *MatchedFeature* nodes with one or more *Feature* nodes. The structure of the *Feature* node is identical to the *Feature* node described for the *FeatureList* format.

1.2. Practical definition of the pLLOQ for spiked human urine samples

Extracted ion chromatograms (EICs) were created using the Bruker Daltonics Data Analysis software, version 3.4 (Bruker Daltonics, Bremen, Germany) for all observed charge states of the 7 added synthetic peptides (VYV, YGGFL, DRVYIHPF, YPFPGPI, YPFPG, GYYPT, YGGWL) and from the five most abundant peptides of digested Carbonic anhydrase (CA). Table S8 (Supplementary material) contains the corresponding m/z values of peptides at the observed charge states. EICs were smoothed using a 3-point Gaussian filter in the retention time dimension, followed by manual integration of the corresponding signal. Signal to noise (S/N) ratios were calculated using the Bruker Daltonics software (Figure S8 in supplementary material). In bioanalytical method validation, a S/N ratio equal or larger than 9 is considered as the lower limit of quantification (LLOQ) for a single compound (1, 2). This definition cannot be applied to our standard mixture used for spiking as it contains many compounds with different LLOQs. For that reason we defined a practical LLOQ

(pLLOQ), which we use throughout this publication. A dilution factor of 200 of the standard stock solution in pooled urine gave an average S/N ratio of 9 (13.5 after leaving out peptide YGGWL (1+)) for the 7 synthetic peptides and for the five most abundant peptides of trypsin-digested CA. We thus defined the spiking of the stock standard solution with a dilution factor of 200 as the practical limit of quantification (pLLOQ). Other dilutions of the standard stock solution are related to the pLLOQ as the reciprocal ratio of the respective dilution factors to the dilution factor of 200.

1.3. Execution of homogenous and heterogeneous data processing workflows

All workflows were executed using two servers equipped with two processors each (8 cores in total) with 16 and 64 GB of random access memory (RAM), respectively, both running Linux (Ubuntu 8.04 LTS). Execution of a few combinations of modules, such as N-M rules feature detection/quantification creating long feature lists, to MZmine or SuperHirn feature alignment/matching required 64 GB of RAM for feature alignment/matching, which is more than normally available on common desktop PCs. Run times per workflow on the described hardware ranged from hours to a couple of days for an analysis of complete dataset urine or porcine CSF) including all possible pairs of spiking concentrations. Processing time can be shortened by using a cluster with more and/or faster CPUs having at least 64 GB of RAM available per processing nodes. Parameters for the original workflows were set according to the experimental setup and guidelines provided with the original homogenous workflows, which includes tailoring parameter settings to the mass resolution and chromatographic peak width as well as to the expected mass and retention time shifts. Limited optimization using manual inspection of the results was performed for the most critical parameters such as retention time, m/z, peak width in retention time and m/z directions using a limited range of values to find the optimal settings. Parameters for all modules, the raw dataset as well as the help file to analyze the raw data with parameters used in this manuscript are available at https://trac.nbic.nl/mscompare/downloads: settings.tar.gz.

1.4. Assignment of features in the spiked human urine dataset that are derived from spiked peptides

A list of features, that are derived from the added peptides (CA digest and synthetic peptides), was assigned based on 4 analyses of samples containing only peptides used for spiking at a 100-fold dilution of the stock solution and analyzing the resulting data with OpenMS, MZmine and SuperHirn. A feature was considered to belong to one of the spiked peptides if it was detected by one of the workflows in at least two separate chromatograms. All features fulfilling these criteria were combined in one set. Features with m/z values below 280 m/z were eliminated, due to a high number of background ions in this region. Furthermore, we considered only features between 30 and 85 min, since no peptides eluted before 30 min and all features eluting later than at 85 min were prone to heavy background and fluctuating ionization due to the high concentration of ACN. This resulted in a list of 1009 putative features corresponding to standard peptides. This list was verified manually by visual inspection of the corresponding EICs in five urine samples spiked at 12.5- and 2000-fold dilution (16 and 0.1 × pLLOQ, respectively) of the standard stock solution. This was followed by deconvolution of isotope clusters to include only monoisotopic features for each peptide-related charge state to result in a final list of 175 identified features. This list corresponded to standard peptides, and constituted our reference list to identify features related to the spiked peptides in feature lists obtained from urine samples. Features identified in the matched feature list of urine samples served as the basis for our scoring function (equation 1) and were used to compare the different data processing workflows. The list of identified standard features is given in Table S6 (Supplementary material).

1.5. Practical definition of the pLLOQ for spiked porcine CSF samples

Extracted ion chromatograms (EICs) were created with Quantitative Analysis of the MassHunter package for all 17 features at the highest spiking level (50 fMol/ul). Each EIC was smoothed using a Gaussian kernel function with a width of 7 after which the features were manually integrated and the S/N values for each feature was calculated (Figure S9 in Supplementary material). Due to the large difference between the median S/N of 24 and average S/N of 61 for these 17 features, we decided to use the median to calculate the pLLOQ, as this would better represent the typical concentration needed to detect these features. Therefore, an S/N of 9 is associated with the 1× LLOQ for a single compound, and would results as approximately 2.5× LLOQ for the same compounds measured with an S/N of 24. The highest spiking level of the CSF sample is equal to 2.5 times the pLLOQ, making the spiked peptide-related features in the porcine CSF dataset more challenging to detect then the spiked peptide-related features in the urine dataset, where the highest spiked concentration level reaches 16 times of the pLLOQ.

1.6. Assignment of features related to spiked peptides in the porcine CSF dataset

A list of spiked peptide-related features was assigned based on 8 analyses of 200 fmol/µl of pure horse hearth cytochrome C tryptic digest solution. This concentration was 4 times higher than the highest spiked concentration. The resulting data were analysed with OpenMS, MZmine and SuperHirn homogenous workflows, and a feature was considered to belong to one of the spiked peptides if it was detected by one of the workflows in at least two separate chromatograms. All features fulfilling these criteria were combined in one set. This list was manually rechecked to remove duplicates and erroneous features using TOPPView from OpenMS, by taking into account the various charge state of one peptide. We carefully checked if the resulting features were detectable in porcine CSF at the highest spiked level of 50 fmol/uL, using the Quantitative Analysis software of the MassHunter package. Features below 200 m/z were eliminated, due to a high number of background ions in this region. Furthermore, we considered only features eluting between 25 and 60 min, since no peptides eluted before 25 min and all features eluting later than at 60 min were prone to heavy background and fluctuating ionization due to the high concentration of acetonitrile. This approach resulted in a final list of 17 features corresponding to horse heart cytochrome C peptides. The list of identified standard features is given in Table S7 (Supplementary material).

1.7. Settings of the score module

We used the score module as discussed in the main text to evaluate the scores for the different combinations of spike levels. The final score is obtained by adding the scores of each concentration pair involved in “Low”, “Medium” and “High” groups of spiked concentration level differences (Table S2 in supplementary material, for human urine dataset and Table S4 for the porcine CSF dataset for an overview over the pLLOQs that were used to construct the different matched feature matrices). The score was applied only to features between 280 – 1500 m/z and elution times ranging from 30 to 85 minutes for the human urine dataset and features between 200 – 2000 m/z and elution times ranging from 25 to 60 minutes for the porcine CSF dataset. Multiple matched feature matrices were constructed based on two spiking concentration levels using 10 LC-MS analyses. Detected and matched features in quantitative feature matrices were ranked according to their *t*-values. Features corresponding to the standard peptides used for spiking were identified based on m/z values and retention times allowing shifts between measured values and the reference list up to ±1 amu for m/z for the human urine dataset and ±0.1 amu for m/z for the porcine CSF dataset and ±1.5 minutes for retention time. This was followed by removing the features corresponding to isotopologues of the spiked compounds in the quantitative matched feature matrix using a tolerance of ±6 sec for the expected retention time and ±0.1 m/z for the expected isotope features for both datasets. Equation 1 was then applied to each sorted feature matrix. The main steps to prepare the raw data for one example of a quantitative matched feature matrix constructed from samples of 1 and 16 times pLLOQ before applying the scoring

function is presented in Figure S1 (Supplementary material). The final scores for the different workflows for "High", "Medium" and "Low" spiking concentration level differences were obtained by summing the scores of each *t*-value sorted quantitative matched feature matrices (see Table S2 and S4, supplementary material; and Figures 3 and 6). Figure 2 exemplifies the results of the scoring function for different constant settings according to equation 1.

1.8. Definition of feature and peak

Peak: A three-dimensional Gaussian-shaped detector signal in single-stage LC-MS data related to a compound. A peak is characterized by the m/z and retention time coordinates of the peak maximum and the quantity of the compound, indicated by the height, area of the extracted ion chromatogram, or peak volume. After a peak is detected by a feature detection/quantification module it is referred to as feature (see below).

Feature: Item in the list obtained after feature detection/quantification in single stage mass spectrometry data. A feature can correspond to a compound-related peak signal, or be an artifact such as local irregular background variation, peak fragment, electronic spike, noise or any data processing artifact. Features are described by similar characteristics as peaks, such as retention time, m/z value and height, area or volume, proportional with the quantities of the original peak.

1.9. References

1. Peters, F., and Maurer, H. (2005) Bioanalytical method validation and its implications for forensic and clinical toxicology — A review. *Validation in Chemical Measurement*, pp. 1-9.
2. Domon, B., and Aebersold, R. (2010) Options and considerations when selecting a quantitative proteomics strategy. *Nat Biotechnol* 28, 710-721.

2. *Supplementary tables (the order of the Tables follows the citation order of the main article Hoekman et al.)*

Concentration (in multiples of pLLOQ)	Dilution factor
16	12.5
8	25
4	50
2	100
1	200
0.5	400
0.1	2000

Table S1. Design of the data set describing the relation between dilution factor and pLLOQ for a standard mixture of peptides that were spiked into a pooled human urine sample (5 LC-MS analyses per spiking level).

Group	Spiking levels of individual samples in terms of pLLOQ	Number of matched feature matrices	Combined concentration pairs expressed in pLLOQ
High	16, 8, 4, 2, 1, 0.5 and 0.1	11	[16, 8]; [16, 4]; [16, 2]; [16, 1]; [16, 0.5]; [16, 0.1]; [8, 4]; [8, 2]; [8, 1]; [8, 0.5]; [8, 0.1]
Medium	4, 2, 1, 0.5 and 0.1	7	[4, 2]; [4, 1]; [4, 0.5]; [4, 0.1]; [2, 1]; [2, 0.5]; [2, 0.1]
Low	1, 0.5 and 0.1	3	[1, 0.5]; [1, 0.1]; [0.5, 0.1];

Table S2. Grouping of spiked human urine samples into “High”, “Medium” and “Low” concentration level differences as used to calculate comparative scores for the different data processing workflows. The second column contains the pLLOQ value of the used concentrations. The third column contains the number of matched feature matrices, where each matched feature matrix is constituted from samples of two different spiking levels based on 5 chromatograms at each level of spiking. The last column lists the pairs of spiking concentrations expressed in pLLOQ that were used to build the matched feature matrices in the corresponding groups. Scores using equation 1 were calculated for each matched feature matrix, and the sum over all matched feature matrices constituted the final scores for one group of spiked concentration level differences.

Concentration (in multiples of pLLOQ)	Concentration (in fMol / ul)
2.5	25
0.5	5
0.25	2.5
0.05	0.5
$5 \cdot 10^{-3}$	0.05
$2.5 \cdot 10^{-3}$	0.025
$5 \cdot 10^{-4}$	$5 \cdot 10^{-3}$

Table S3. Design of the data set describing the relation between the concentration in fMol/uL and pLLOQ for the peptides resulting from the trypsin digestion of horse hearth cytochrome C that were used for spiking into a single CSF sample (5 LC-MS analyses per spiking level). At a concentration of 0.5 fMol/uL none of the 17 features are detectable at a S/N of 10 or more.

Group	Spiking levels of individual samples in terms of pLLOQ	Number of matched feature matrices	Combined concentration pairs expressed in pLLOQ
High	2.5, 0.5, 0.25, 0.05, $5 \cdot 10^{-3}$, $2.5 \cdot 10^{-3}$ and $5 \cdot 10^{-4}$	6	[2.5, 0.5]; [2.5, 0.25]; [2.5, 0.05]; [2.5, 5^{e-3}]; [2.5, 2.5^{e-3}]; [2.5, 5^{e-4}];
Low	0.5, 0.25, 0.05, $5 \cdot 10^{-3}$, $2.5 \cdot 10^{-3}$ and $5 \cdot 10^{-4}$	5	[0.5, 0.25]; [0.5, 0.05]; [0.5, 5^{e-3}]; [0.5, 2.5^{e-3}]; [0.5, 5^{e-4}];

Table S4. Grouping of the spiked CSF samples into “High” and “Low” concentration level differences used to calculate comparative scores for the different data processing workflows. The second column contains the pLLOQ value corresponding to the spiked concentrations. The third column contains the number of matched feature matrices. Each constituted from samples of two different spiking levels based on 5 LC-MS chromatograms at each level of spiking. The last column lists the pairs of spiking concentrations expressed in pLLOQ that were used to prepare the matched feature matrices in the corresponding groups. Scores were calculated using equation 1 for each matched feature matrix, and the sum over all matched feature matrices constituted the final scores for one group of spiked concentration level differences.

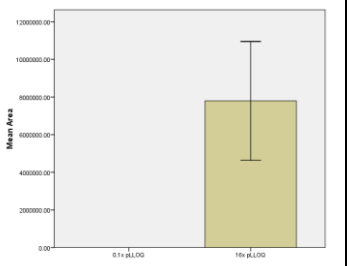
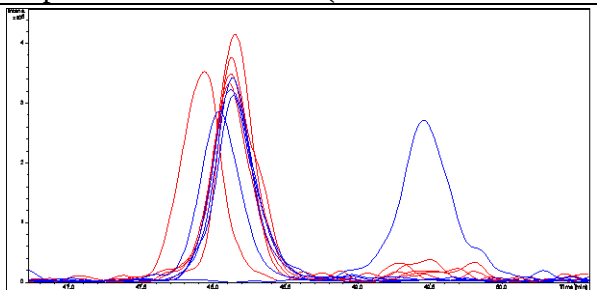
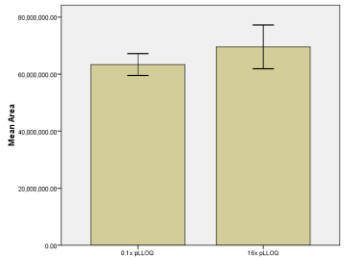
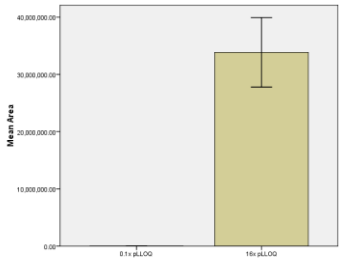
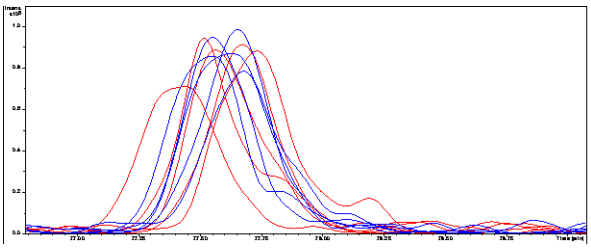
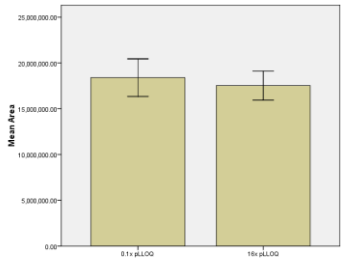
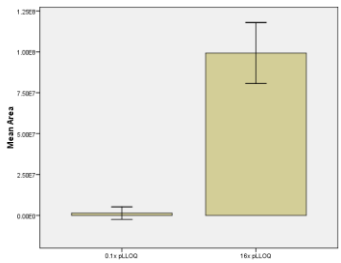
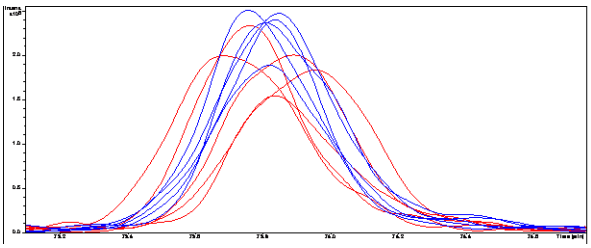
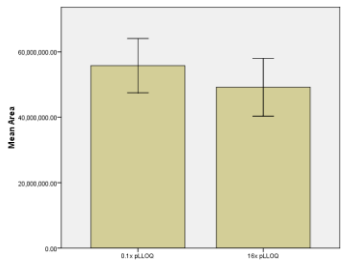
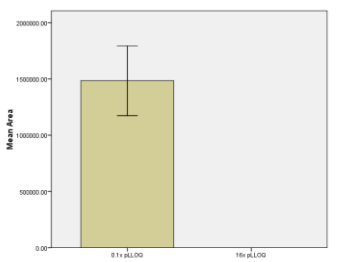
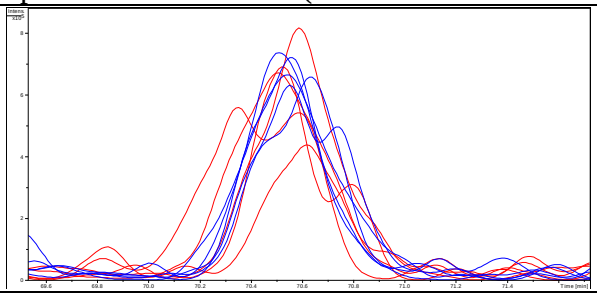
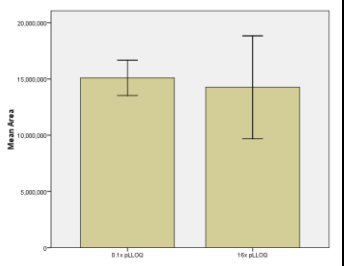
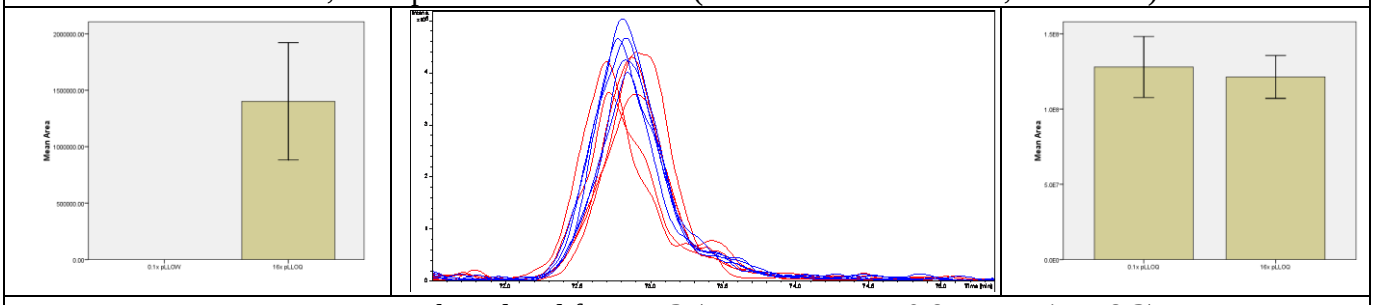
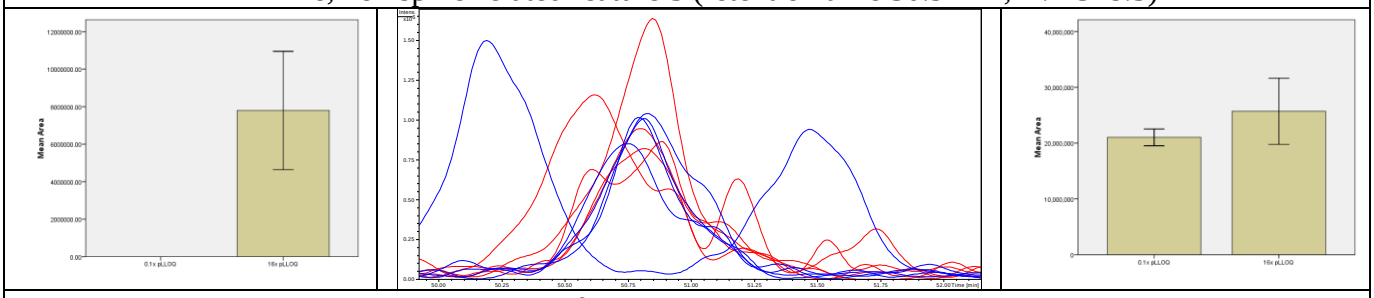
Bar graph of feature in 5 samples of 0.1× and 16× pLLOQ obtained from the matched feature matrix.	EICs of feature obtained with Bruker Daltonics DataAnalysis software. Blue lines correspond to traces from samples spiked at 0.1× pLLOQ and red lines correspond to samples spiked at 16× pLLOQ.	Bar graph of feature in 5 samples of 0.1× and 16× pLLOQ obtained with manual integration of EICs.
SuperHirn, non spike-related feature 1 (retention time 48 min, m/z 547.1)		
		
SuperHirn, non spike-related feature 2 (retention time 77.6 min, m/z 651.2)		
		
SuperHirn, non spike-related feature 3 (retention time 73.6 min, m/z 439.2)		
		
MZmine, non spike-related feature 1 (retention time 70.5 min, m/z 365.1)		
		

Table S5. First part. See legend at the end of the third part.

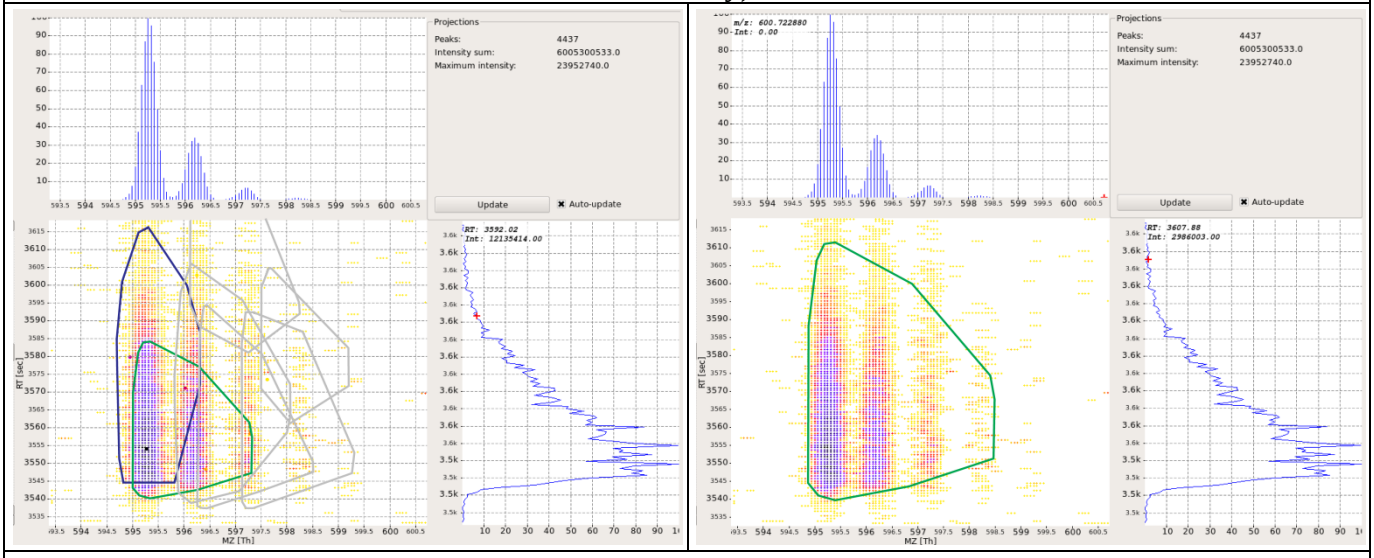
MZmine, non spike-related feature 2 (retention time 73.4 min, m/z 460.1)



MZmine, non spike-related feature 3 (retention time 50.9 min, m/z 578.3)



OpenMS, non spike-related feature 1 (retention time 59.6 min, m/z 595.0)
(left graph shows feature detection by OpenMS, right graph shows the correct feature selected manually)



OpenMS, non spike-related feature 2 (retention time 59.2 min, m/z 449.0)

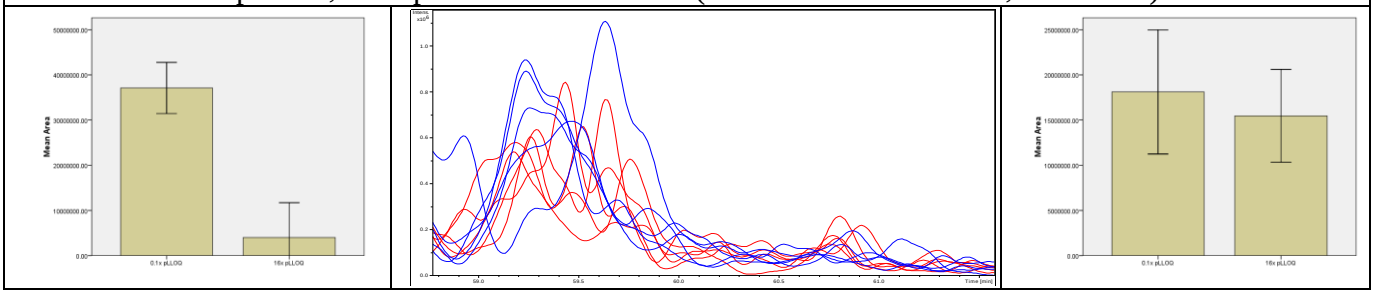


Table S5. Second part. See legend at the end of the third part.

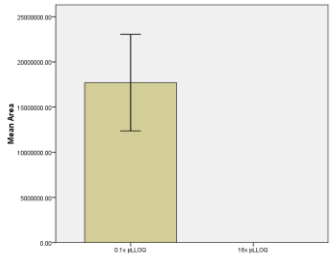
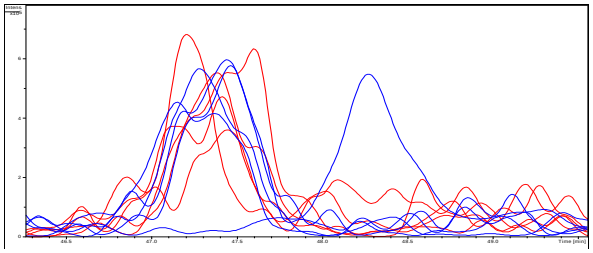
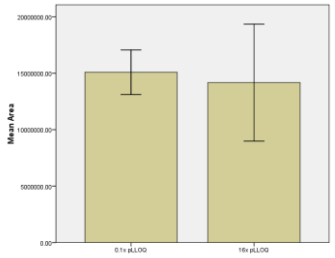
OpenMS, non spike-related feature 3 (retention time 47.2 min, m/z 381.1)		
		

Table S5. The 3 highest discriminative (based on *t*-score) non spiked peptides-related features resulted in as data processing errors for the homogenous workflows of OpenMS, MZmine and SuperHirn obtained with the urine sample spiked at 16× pLLOQ level and blank spiked at 0.1× pLLOQ level). Majority of the data processing errors occur due to splitting up of one peak at feature detection/quantification step, which is incorrectly matched to the right features of the samples of the other spiking level. The highest discriminative feature obtained with homogenous OpenMS pipeline considered as a data processing error is due the double detection of the one spiked peptide-related chromatographic peak. The double detection of the spiked peptide resulted in two highly significant features of which only one can be considered to be related to the peaks of the spiked peptide.

no	rt (mins)	m/z
1	31.5	506.65
2	34.5	344.94
3	34.5	338.46
4	35.25	468.53
5	35.25	409.28
6	35.75	501.35
7	37.4	379.90
8	38.8	322.02
9	39.8	291.48
10	39.8	600.16
11	39.8	300.55
12	39.8	533.71
13	40.7	490.19
14	41.0	541.17
15	41.2	518.71
16	42.0	721.84
17	43.3	353.23
18	43.0	597.69
19	41.8	443.09
20	41.8	569.24
21	42.6	414.61
22	42.8	648.33
23	42.8	324.61
24	42.8	493.19
25	43.0	985.31
26	43.4	353.13
27	43.4	424.58
28	43.4	401.5
29	43.3	321.5
30	43.3	723.75
31	43.5	509.28
32	44.7	491.09
33	44.2	538.14
34	44.6	524.04
35	44.5	485.82
36	44.7	504.57
37	45.3	351.46
38	45.7	586.33
39	46.3	520.14
40	46.0	504.42
41	46.7	502.48
42	46.5	418.09
43	46.5	500.99
44	46.7	653.28
45	46.7	673.82
46	46.7	449.51
47	46.7	558.53
48	47.0	529.75
49	47.2	383.46
50	47.2	420.53

No	rt (mins)	m/z
51	47.8	411.67
52	47.8	338.67
53	47.8	439.47
54	48.2	451.14
55	48.3	624.27
56	48.7	543.56
57	49.0	552.27
58	49.0	375.26
59	49.1	356.50
60	49.5	290.55
61	49.7	580.02
62	50.2	453.06
63	49.0	350.58
64	50.1	520.60
65	50.2	593.73
66	50.2	566.73
67	50.3	571.54
68	50.5	398.51
69	52.6	552.91
70	50.7	355.48
71	51.0	523.75
72	51.0	569.69
73	50.7	349.40
74	50.7	513.21
75	50.9	383.96
76	51.0	575.80
77	51.0	604.23
78	51.5	486.99
79	51.7	396.18
80	51.9	527.89
81	52.0	600.20
82	52.1	429.67
83	52.3	515.60
84	52.5	552.89
85	52.9	546.85
86	53.0	681.94
87	53.1	701.00
88	53.7	578.28
89	54.0	626.17
90	54.0	459.48
91	54.3	500.97
92	54.7	412.35
93	55.2	590.12
94	55.3	606.76
95	55.3	725.36
96	55.7	555.98
97	55.7	650.42
98	56.3	611.72
99	56.4	683.43
100	56.4	596.68

No	rt (mins)	m/z
101	57.4	470.52
102	56.4	470.52
103	56.4	894.41
104	56.4	615.60
105	56.6	451.54
106	56.8	506.56
107	57.2	489.56
108	57.0	526.19
109	57.2	371.71
110	57.3	570.74
111	57.7	710.52
112	57.7	583.23
113	58.0	599.11
114	58.2	647.82
115	58.5	716.53
116	59.0	611.76
117	59.0	652.04
118	59.0	780.86
119	59.3	595.23
120	59.3	298.02
121	59.7	595.02
122	60.7	511.12
123	60.7	667.90
124	60.7	445.36
125	60.7	696.09
126	61.3	705.55
127	61.4	528.21
128	61.5	724.97
129	62.0	561.91
130	62.7	401.54
131	62.8	474.28
132	63.0	757.77
133	63.7	395.66
134	63.7	790.17
135	63.5	693.31
136	64.5	578.27
137	64.5	637.65
138	64.6	549.80
139	66.3	966.90
140	66.3	645.19
141	66.7	723.77
142	66.9	765.34
143	67.9	733.38
144	67.9	554.79
145	67.9	752.34
146	67.9	550.35
147	68.0	582.38
148	68.0	776.03
149	68.3	742.97
150	69.0	771.34

No	rt (mins)	m/z
151	69.1	578.50
152	69.0	652.26
153	69.3	738.03
154	69.3	728.12
155	69.3	621.19
156	69.5	678.08
157	69.7	874.65
158	69.7	583.23
159	70.2	569.55
160	70.3	759.03
161	70.3	777.95
162	71.7	657.24
163	72.0	810.82
164	72.3	839.02
165	74.7	671.28
166	74.7	808.78
167	74.8	938.61
168	76.7	710.71
169	77.0	751.68
170	77.1	770.32
171	77.2	325.02
172	77.7	675.27
173	78.7	694.33
174	79.0	801.54
175	79.7	740.00

Table S6. Retention times and mass to charge ratios of the monoisotopic standard peaks derived from spiked peptides (7 synthetic peptides and peptides of trypsin-digested CA) in spiked human urine. Ions with different charge states are considered as separate features.

no	rt (mins)	m/z	charge state
1	31.25	735.89	2
2	31.40	490.94	3
3	31.40	728.88	2
4	38.90	390.26	2
5	39.40	390.24	3
6	38.60	217.16	1
7	39.40	337.55	3
8	39.10	534.33	2
9	39.10	584.85	2
10	41.50	450.94	3
11	43.40	598.85	2
12	43.90	482.81	2
13	43.90	964.59	1
14	45.30	541.97	3
15	48.25	734.40	2
16	48.70	748.40	2
17	48.60	499.27	2

Table S7. Retention times, mass to charge ratios and charge states of the monoisotopic standard peaks derived from spiked peptides (tryptic digest of horse hearth Cytochrom C) in porcine CSF samples. Peptides with different charge states are considered as separate features.

No.	Peptide (sequence)	m/z values of EICs[a] ($[M+H]^{1+}$, $[M+2H]^{2+}$, $[M+3H]^{3+}$, $[M+4H]^{4+}$)
1	VYV (synthetic)	380.2
2	GYYPT (synthetic)	600.3, 300.7
3	YPFPG (synthetic)	580.3
4	DRVYIHPF (synthetic)	523.8, 349.5
5	YGGFL (synthetic)	556.3
6	YGGWL (synthetic)	595.3
7	YPFPGPI (synthetic)	790.4, 395.7
8	LVQFHFHWGSSDDQGSEHTVDR (CA)	645.2
9	AVVQDPALKPLALVYGEATSR (CA)	733.7
10	RMVNNGHSFNVEYDDSQDK (CA)	759.4
11	VLDALDSIK (CA)	487.0
12	DGPLTGTYR (CA)	490.2

Table S8. m/z values of the 7 synthetic peptides and the five most abundant peptides of trypsin-digested CA used for spiking human urine sample at different observed charge states.

Supplementary Figures
(the order of the figures follows the citation order of the main article Hoekman et al.)

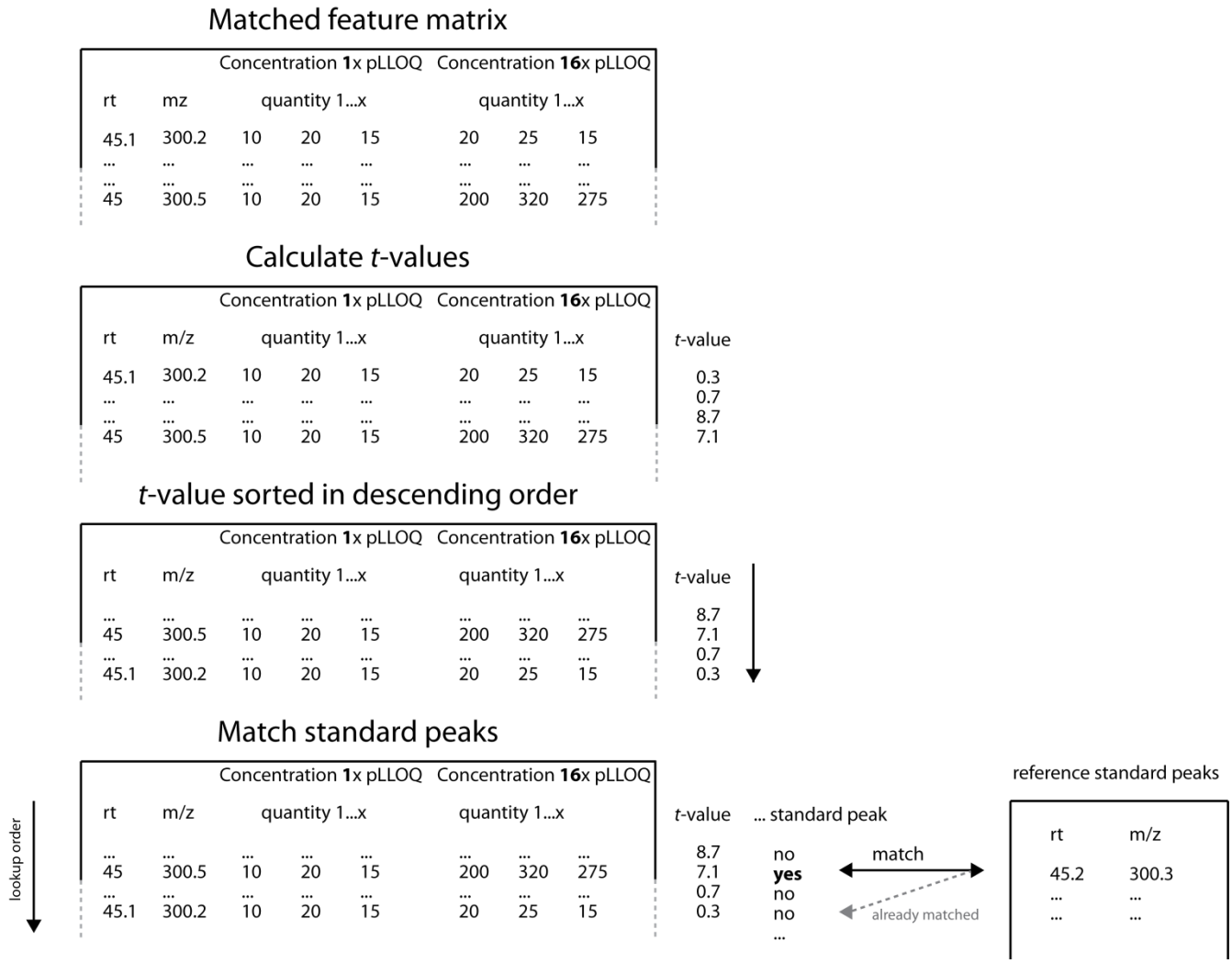


Figure S1. Steps of identifying features corresponding to the spiked peptides present in the matched feature matrix constructed using urine samples that were spiked at 1 and 16-times pLLOQ. Each row corresponds to one feature characterized by its retention time (rt) and mass to charge ratio (m/z). In the unsorted matched feature matrix two features are shown, one corresponding to a spiked-peptide, and the other to another feature. First *t*-value is calculated using independent *t*-tests between samples with 1× and 16× pLLOQ and all features are sorted according to the obtained *t*-values in descending order. This is followed by the identification of spiked-peptides related features and the score for the matched feature matrix is calculated using equation 1. Figure 2 in the main article Hoekman *et al.* shows the operational mechanism of the scoring function.

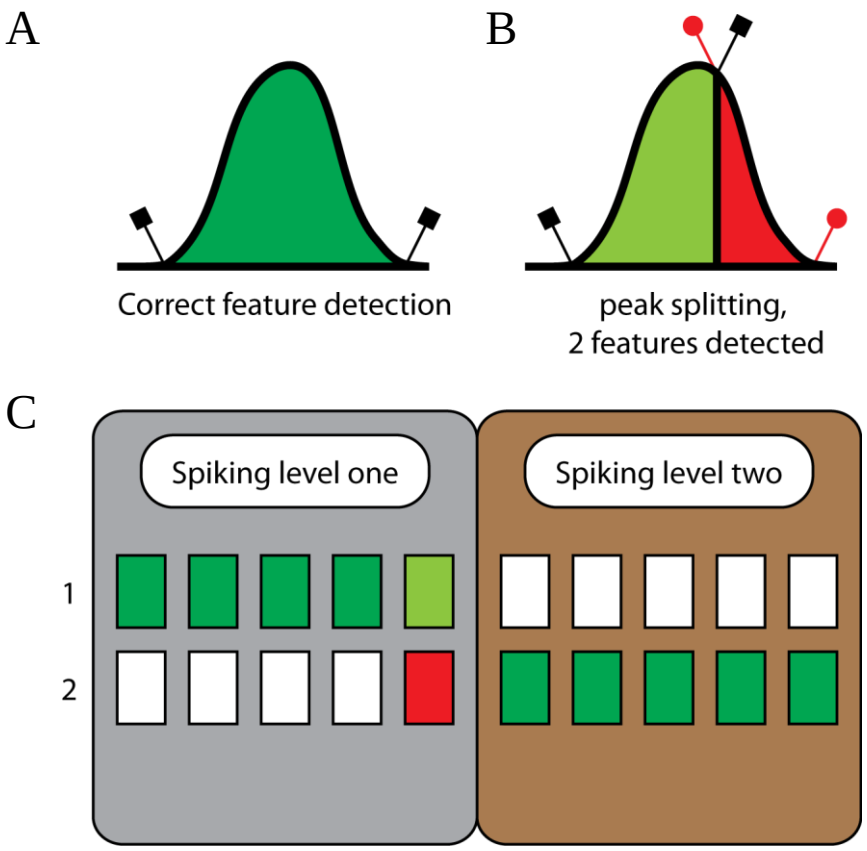


Figure S2. One of the most often occurring error in quantitative matched feature matrix results from the inaccurate combination of information obtained by the feature detection/quantification and feature alignment/matching modules. One peak present in all of the 10 analysed LC-MS chromatograms is identified as two features in one of the chromatograms at one concentration level (B) due to a feature detection error, while the same peak in all other chromatograms is detected and quantified accurately (A). One of the two features resulting from the split of the peak is matched to features corresponding to the same peak in other chromatograms of the same spiking level, while the feature corresponding to the other part of the split peak is matched to the features corresponding to the same peak in chromatograms of the other spiking level (C). This error results in one matched feature (row 1 in C) highly discriminative between the two spiking level, although all the matched features are derived from the same peak present in the same quantity in all chromatograms.

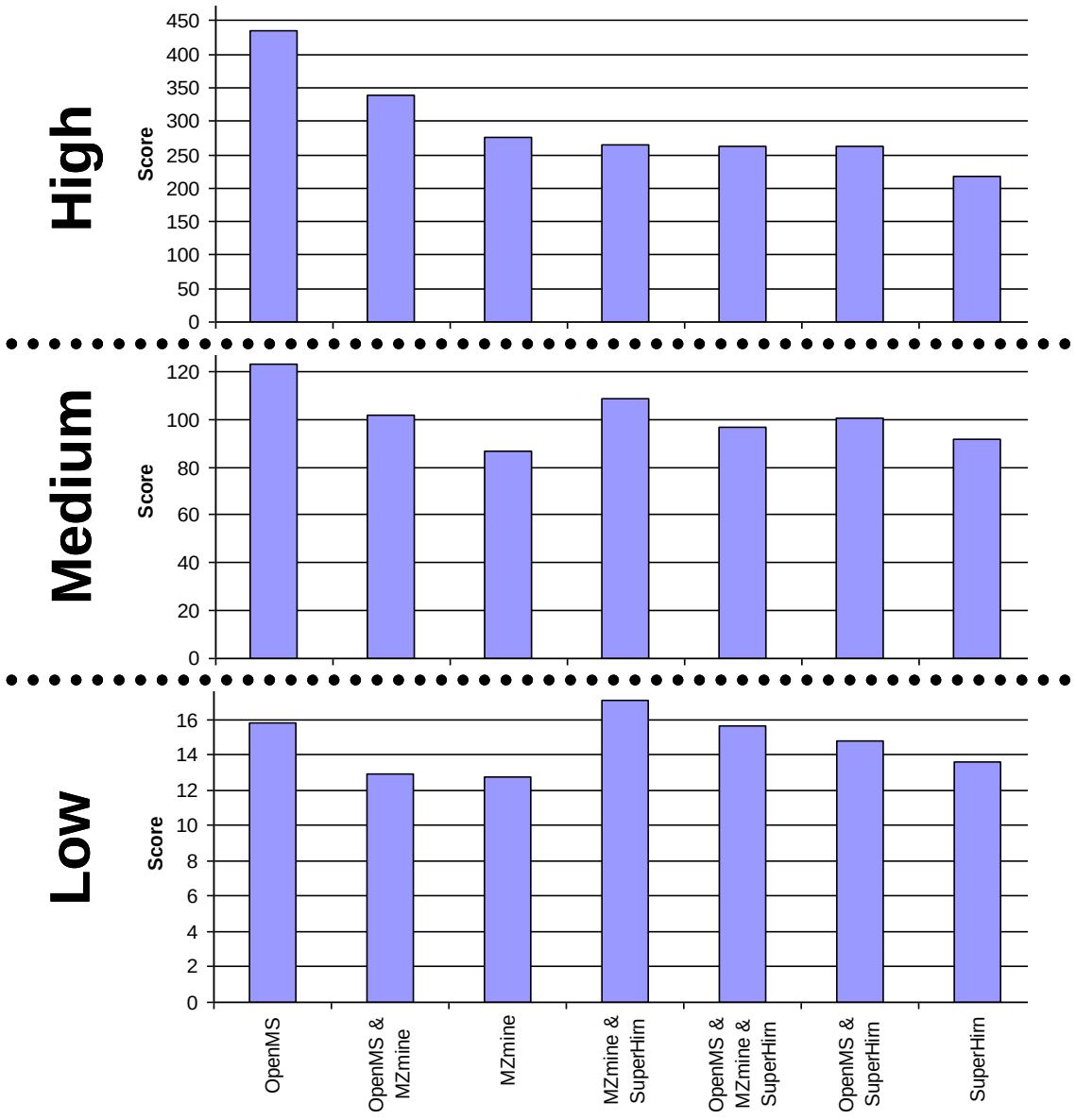


Figure S3. Comparison of the performance for the homogenous workflows of OpenMS, SuperHirn and MZmine and all feature matrices merged with union combination obtained from spiked human urine samples. The OpenMS homogenous workflow shows the best performance for low and medium spiked concentration differences, while it shows good performance at low spiked concentration differences. Merging the feature matrices of MZmine and SuperHirn results in higher scores than the scores obtained for feature matrices that were obtained with the individual homogenous workflows.

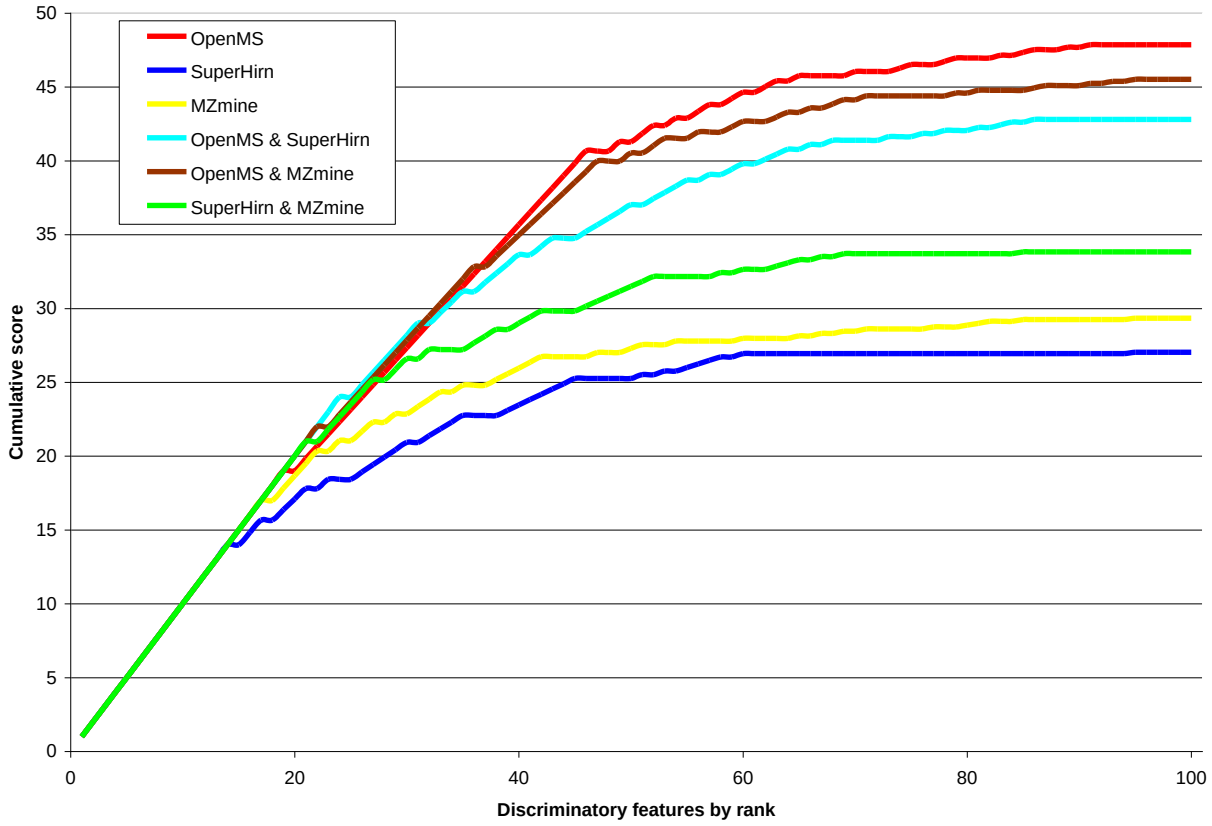


Figure S4. Overview of the score evaluation as a function of the rank of the most discriminating features for the three homogenous workflows and for three feature matrices that were obtained after merging with union combination. Feature matrices were obtained using human urine samples spiked at 16-times pLLOQ vs. the blank (0.1-times pLLOQ).

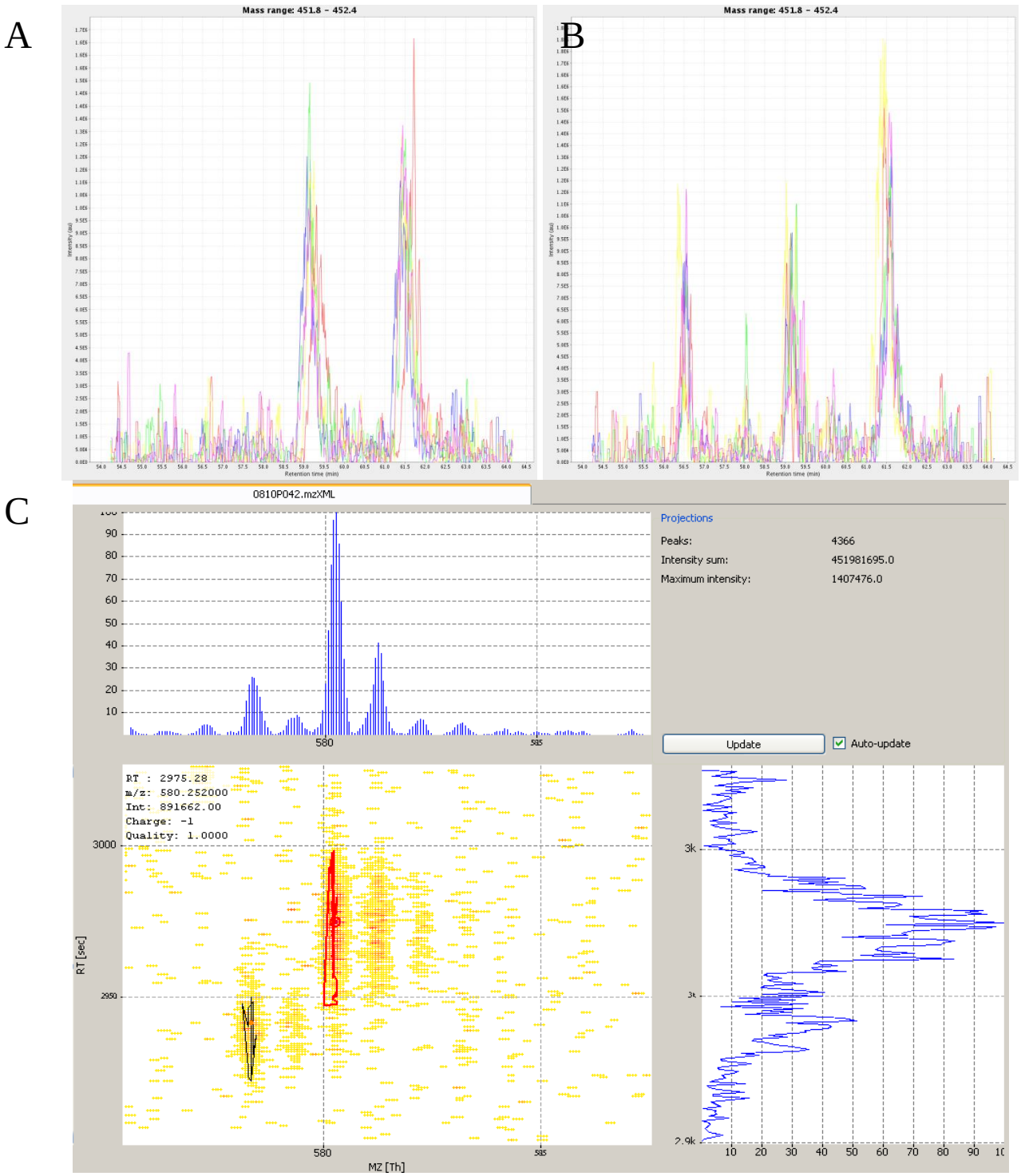


Figure S5. Visualization possibilities of the msCompare framework. A and B show close view of compound peaks in a set of overlaid extracted ion chromatograms obtained from raw human urine LC-MS chromatograms of two spiking groups (A, 2× pLLOQ and B, 16× pLLOQ). EIC of individual chromatograms are colored differently. C shows an overlay of a piece of raw LC-MS data with two features in red and black contours detected with the MZmine feature detection/quantification module. The visualization is performed in the TOPPView visualization program, which is part of the OpenMS framework. msCompare allows this type of visual quality control of the detected features obtained with all feature detection/quantification modules integrated in msCompare.

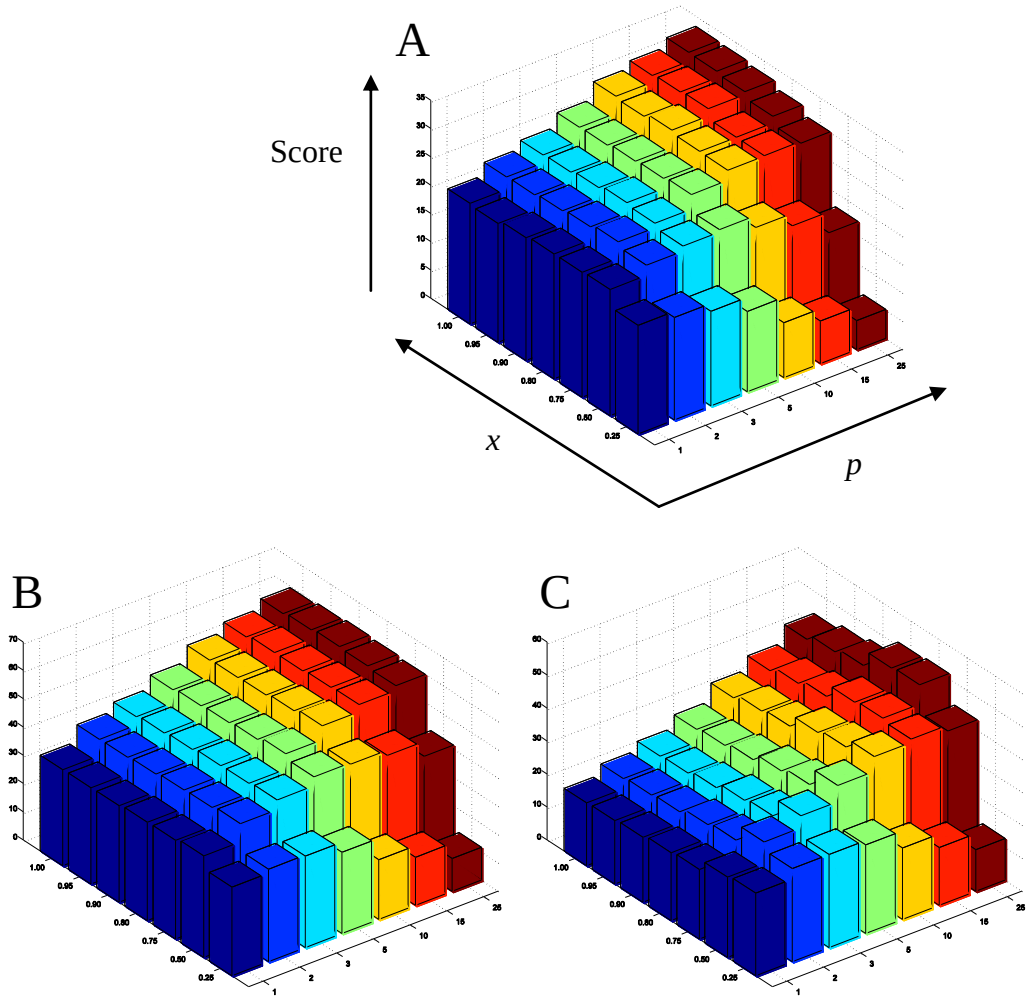


Figure S6. Optimization of the score function parameters *p* and *x* using a set of human urine samples spiked at 16× and 0.1× pLLOQ obtained with the homogenous workflows of MZmine (A), OpenMS (B) and SuperHirn (C). The parameters *x* and *p* were varied over the range from 0.25 to 1 for the parameter *x* and from 1 to 25 for the parameter *p*. The similar plots obtained with three homogenous workflows show that the score calculated with equation 1 increases with increasing *p*-value with exception for low values of *x*. A relatively stable score region was observed for *x* larger or equal than 0.75, and for *p* smaller or equal than 25. Parameters selected within this region are therefore suitable for the relative quantification assessment of the different data processing workflows.

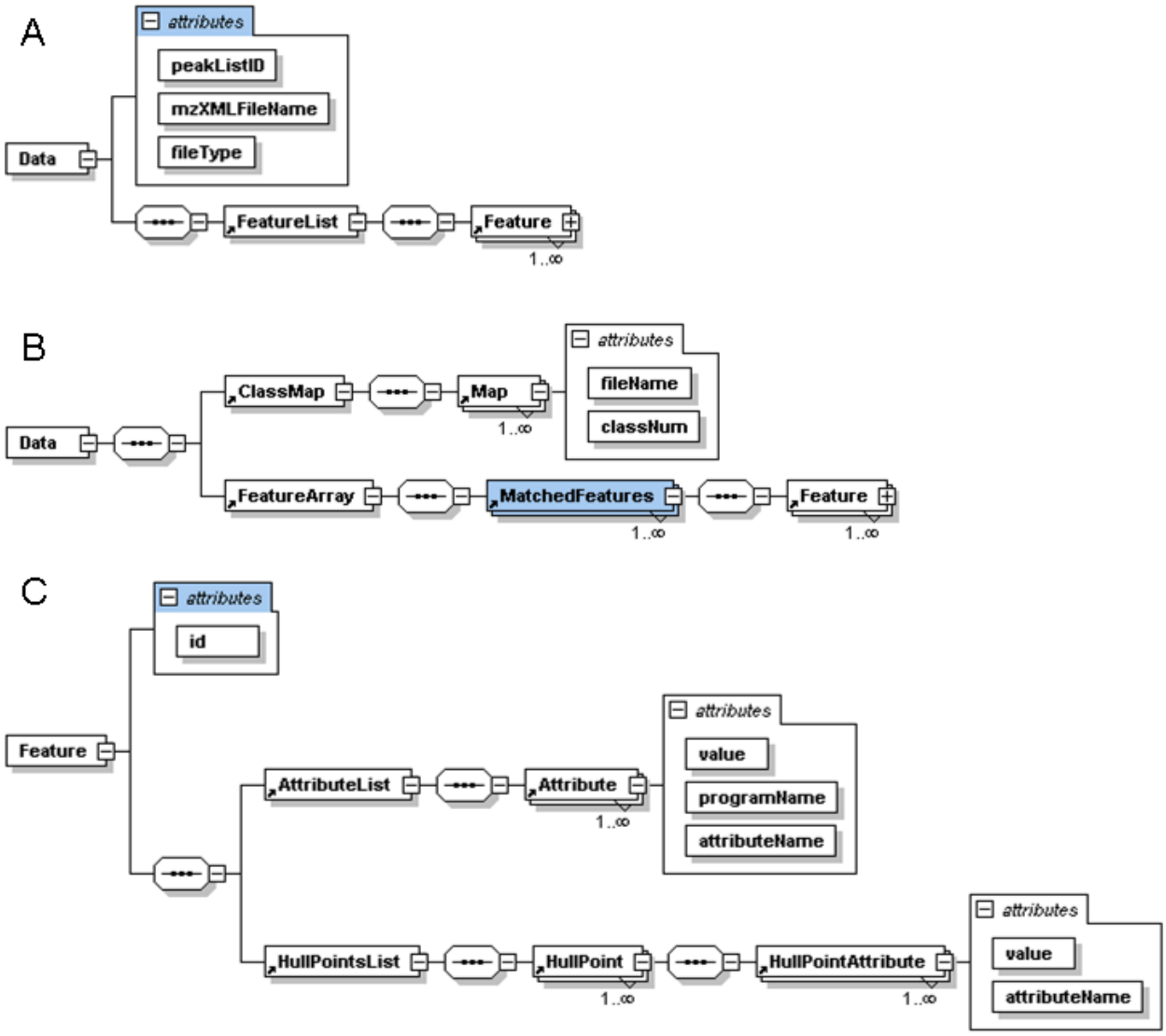


Figure S7. XML-scheme for the two XML-formats used by msCompare: A) XML-scheme of the FeatureList-XML format, B) XML-scheme of the FeatureMatrix-XML format. The expansion of the feature node (same for both formats) is shown under C).

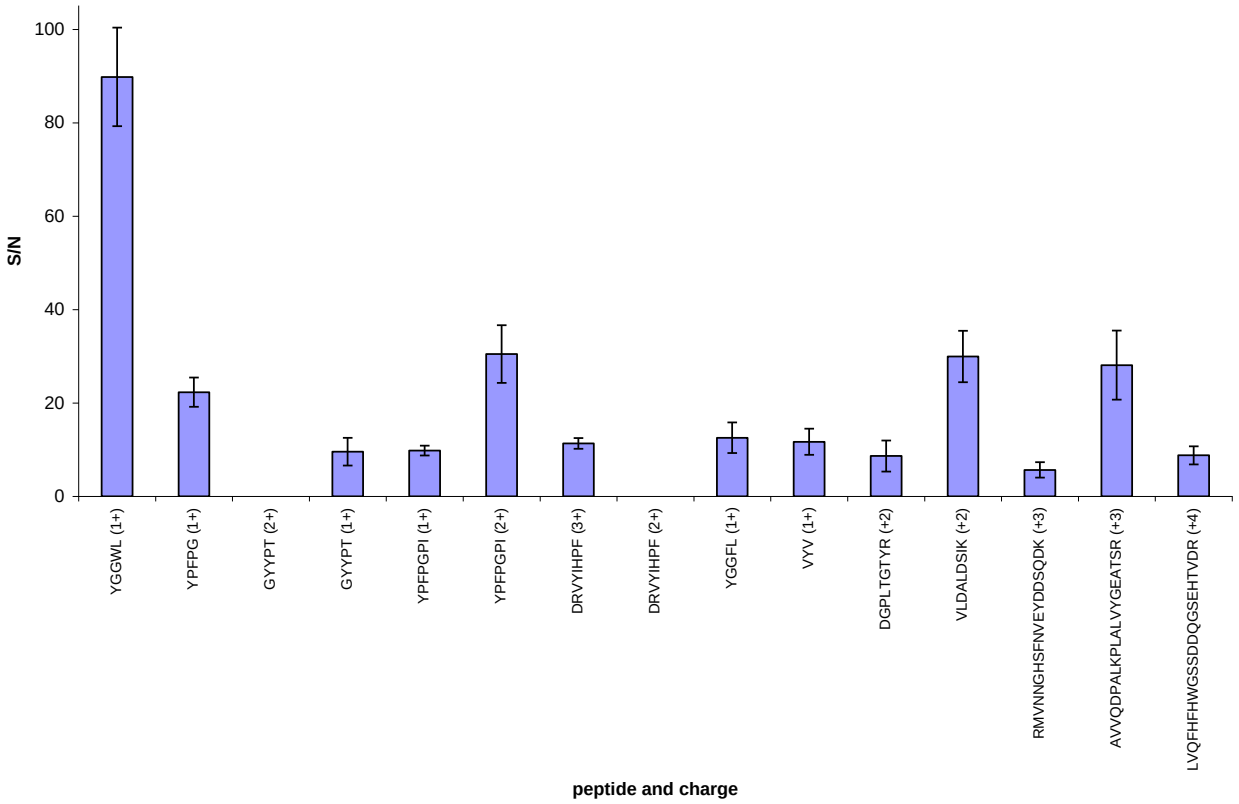


Figure S8. Signal-to-noise ratios (S/N) of peaks derived from 7 synthetic peptides and the five most abundant peptides of trypsin-digested CA that were spiked into a human pooled urine sample at a dilution factor of 200. Peaks related to GYYPT(2+) and DRVYIHPF(2+) were not detectable at this spiking level. Although the S/N ratios show large fluctuations, most of the peaks have S/N ratios of 9 or higher. Therefore, we defined this spiking level to be equivalent to one times pLLOQ.

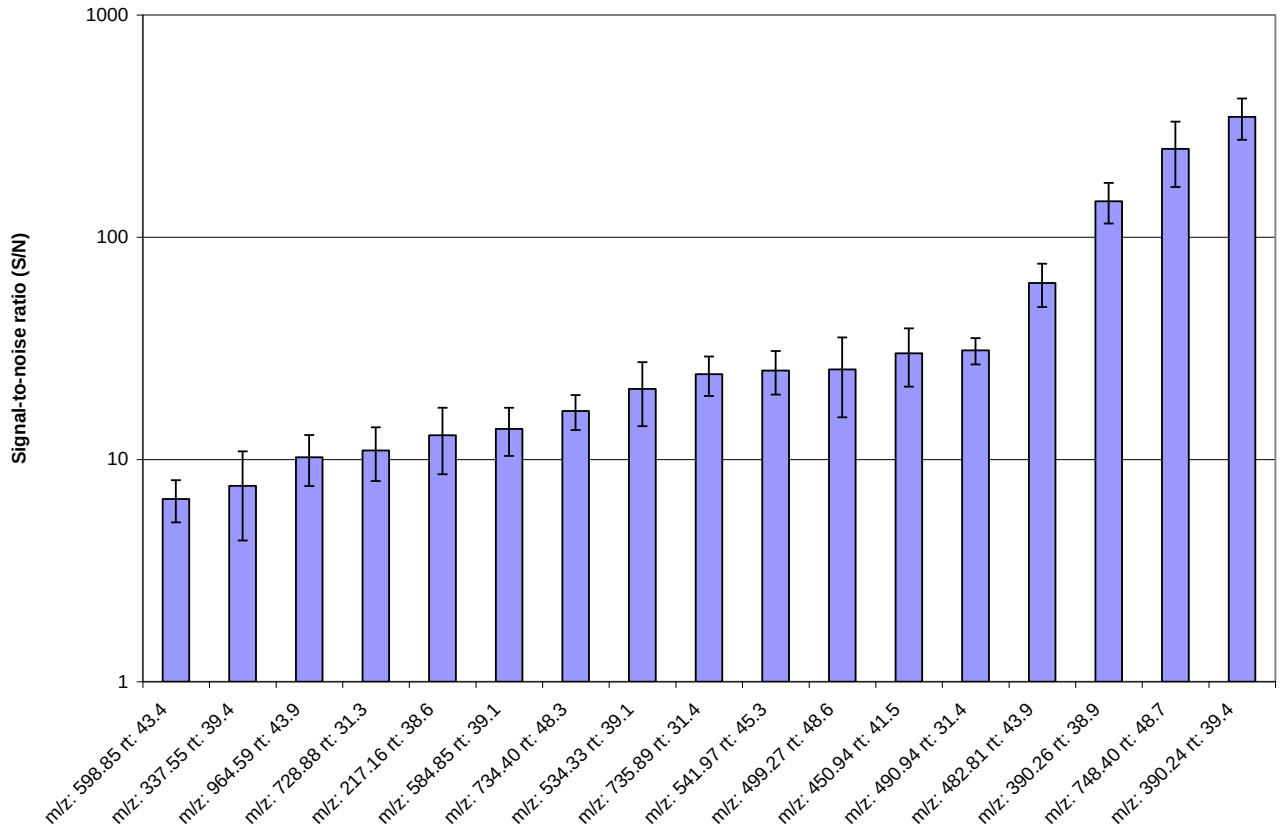


Figure S9. S/N of the peaks derived from trypsin-digested horse hearth cytochrome C spiked into porcine CSF sample at a final concentration of 25 fMol/uL. The mean S/N value of 61 of the spiked-in peptides is considerably different from the median, which is 24 S/N. The median value of 24 S/N is defined to be equivalent to 2.5 times pLLOQ.